

REPRESENT FIRST MOLECULE AS A SET  
OF ATOMS WITH ASSOCIATED SCALAR  
DESCRIPTORS DERIVED FROM INTERATOMIC  
DISTANCES IN SAID FIRST MOLECULE 12

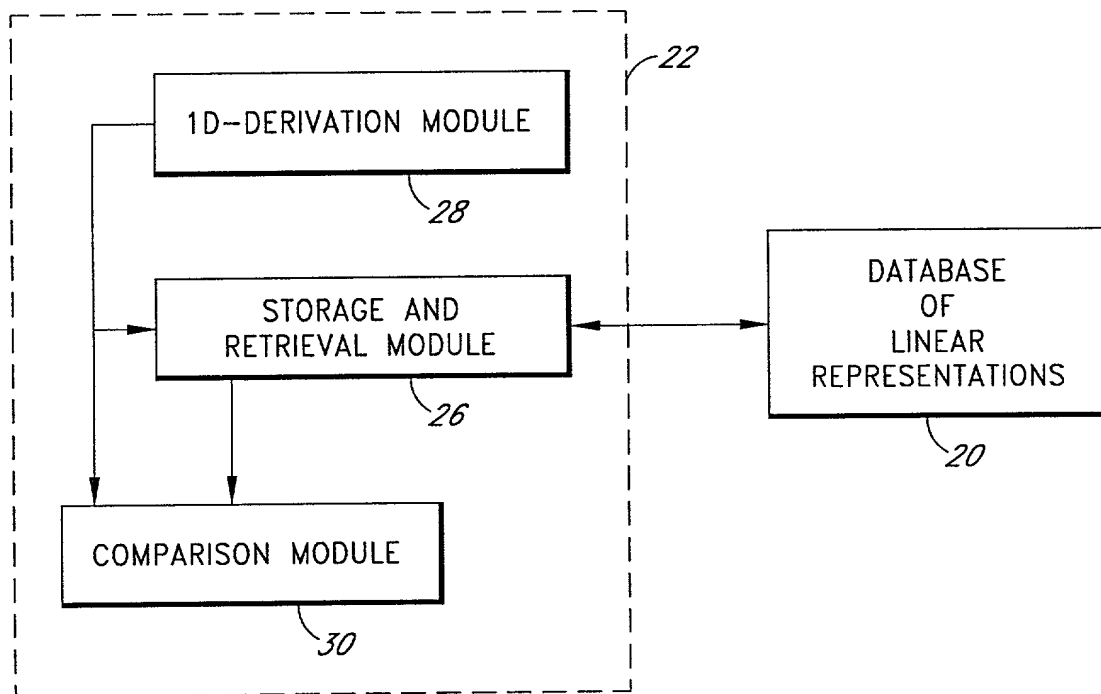


REPRESENT SECOND MOLECULE AS A SET  
OF ATOMS WITH ASSOCIATED SCALAR  
DESCRIPTORS DERIVED FROM INTERATOMIC  
DISTANCES IN SAID SECOND MOLECULE 14



ASSESS MOLECULAR SIMILARITY  
BY COMPARING THE MOLECULAR  
REPRESENTATIONS 16

FIG. 1



**FIG. 2**



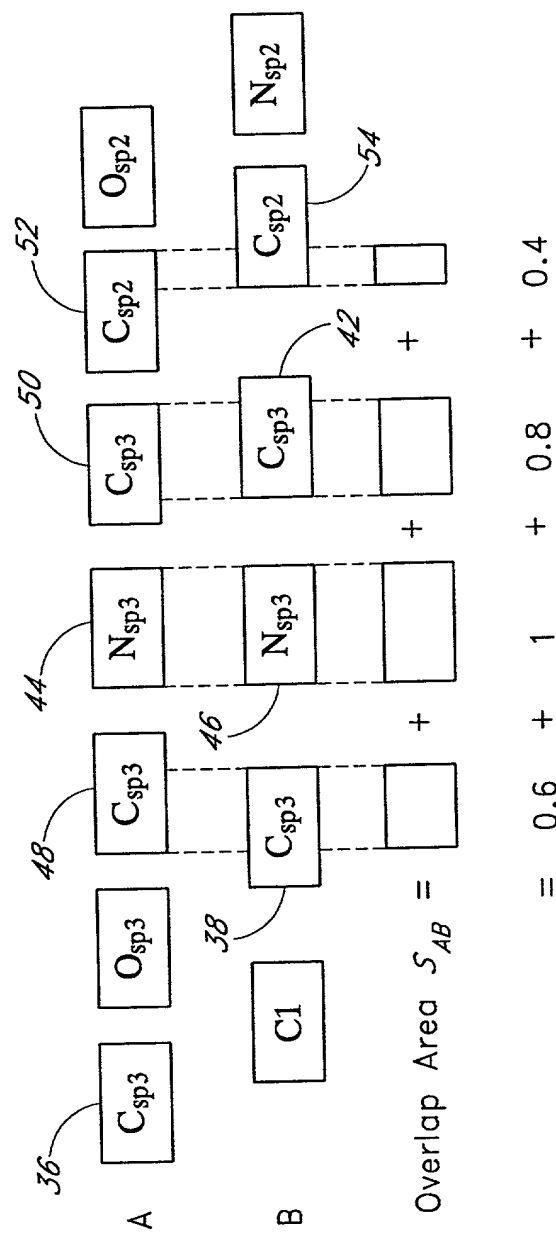


FIG. 3B

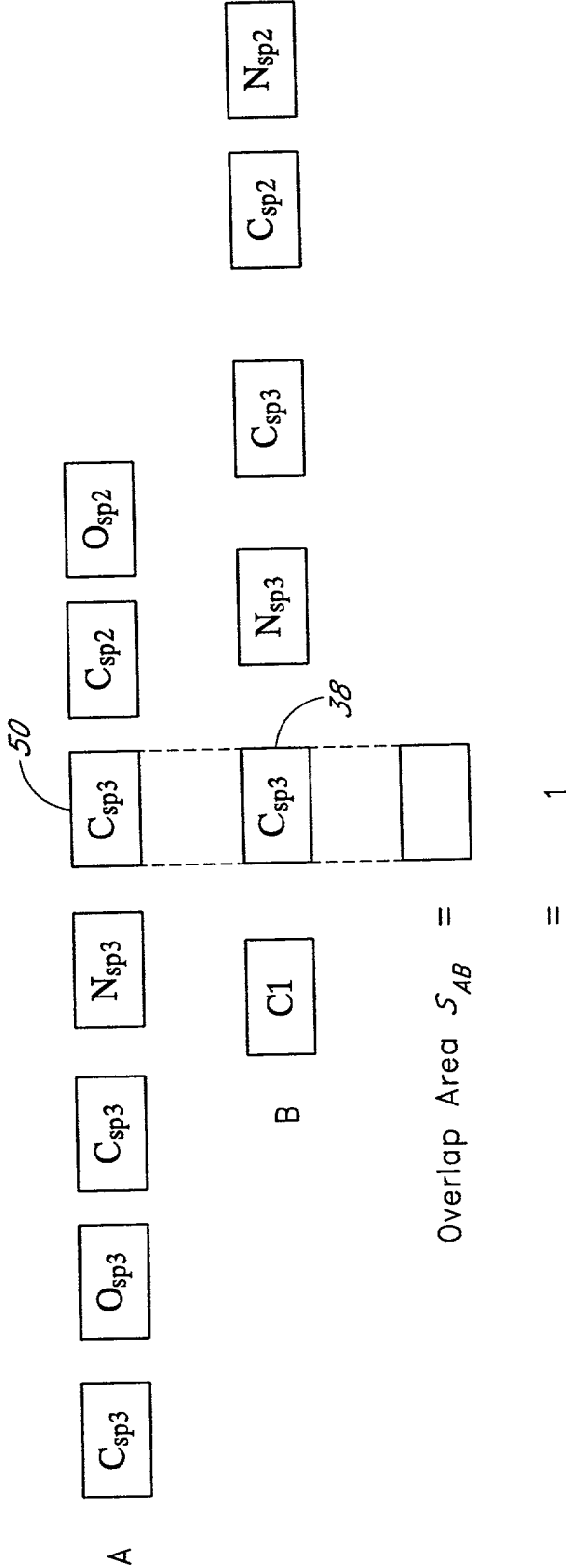


FIG. 3C

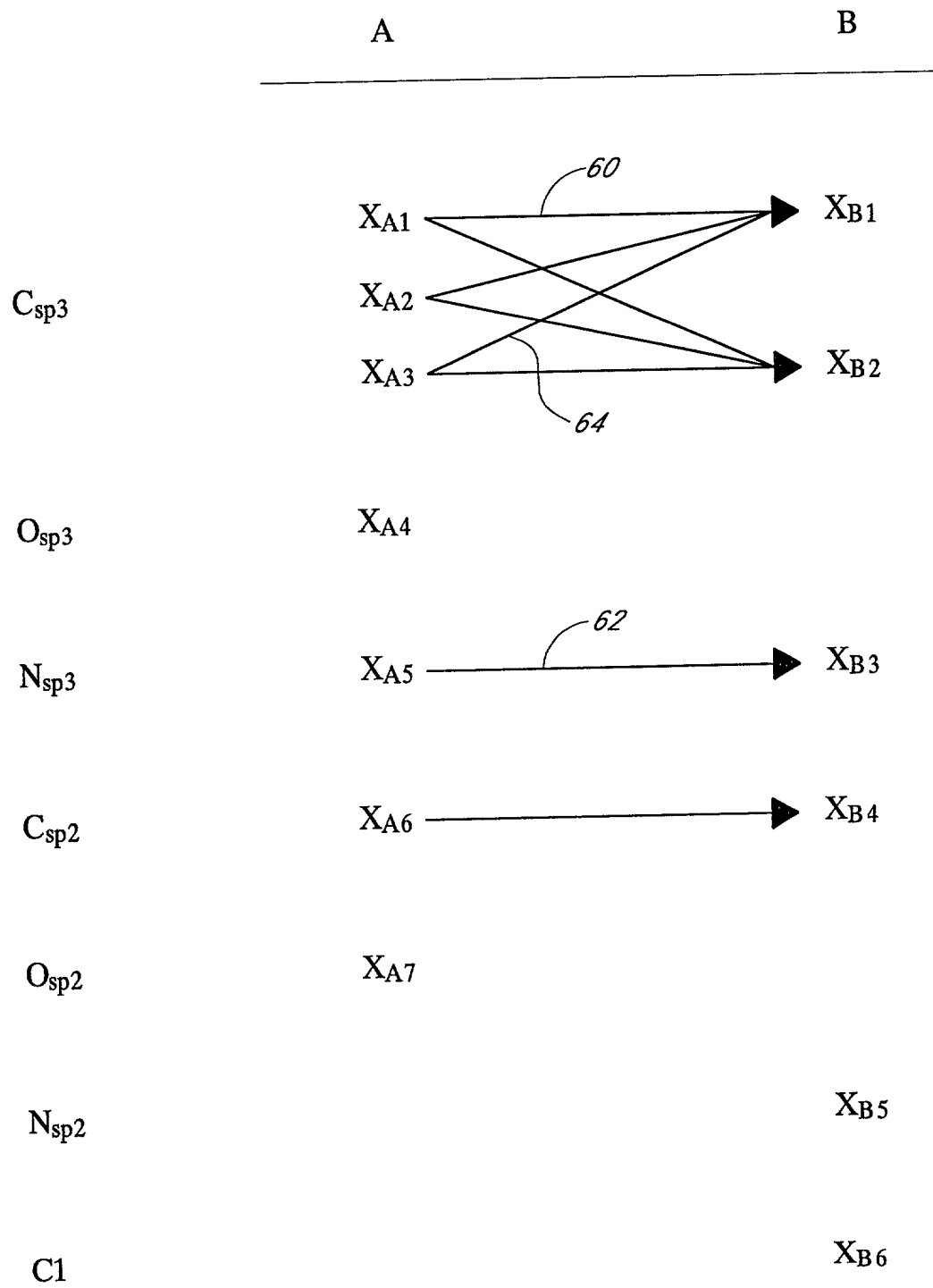


FIG. 4

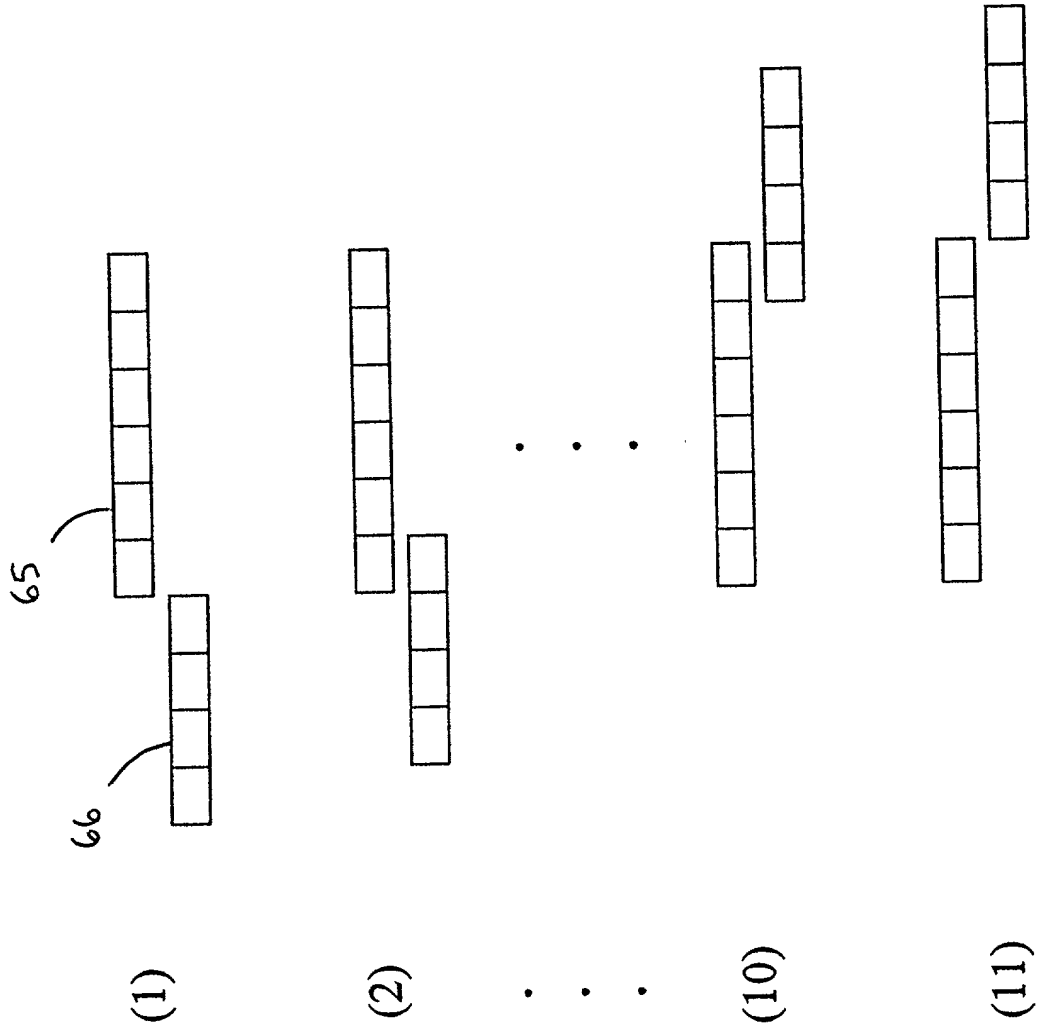


FIG. 5

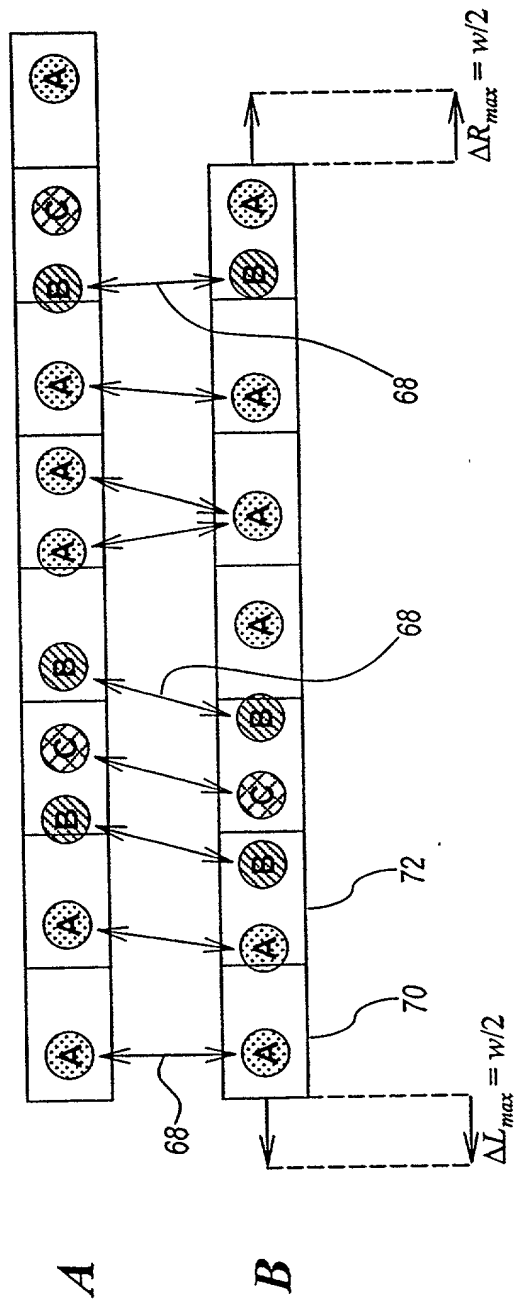
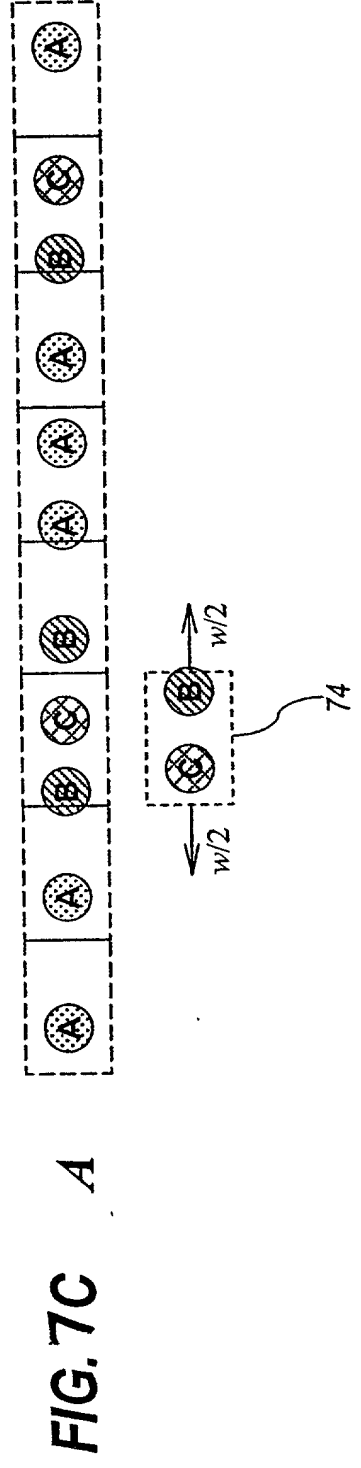
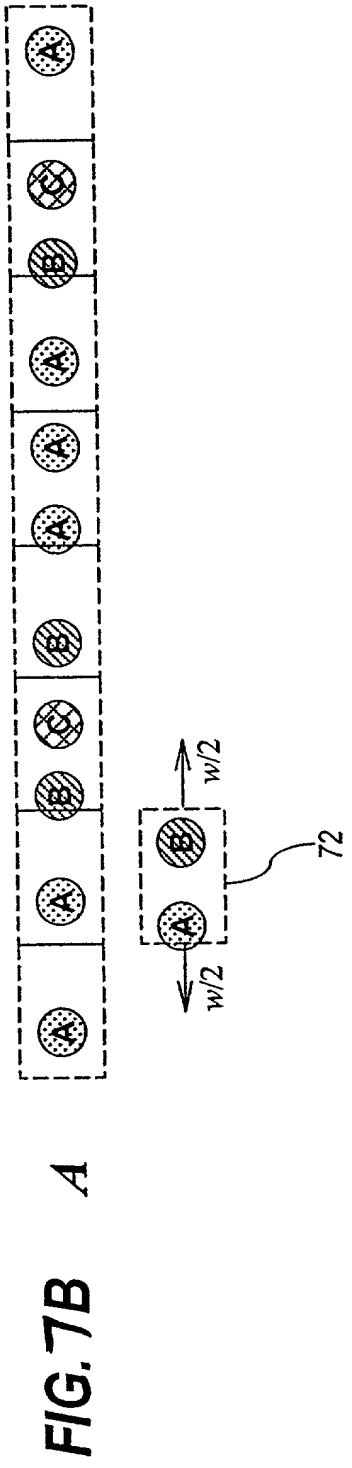
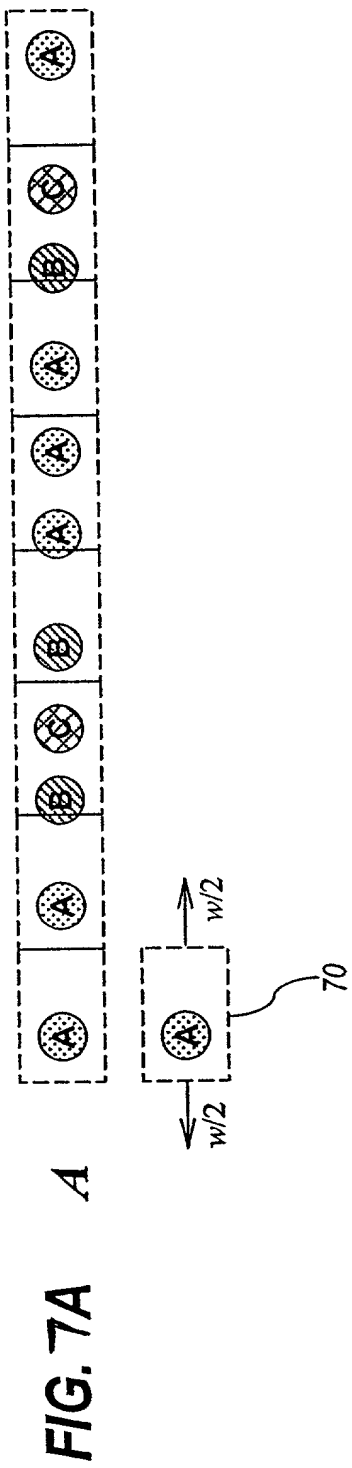
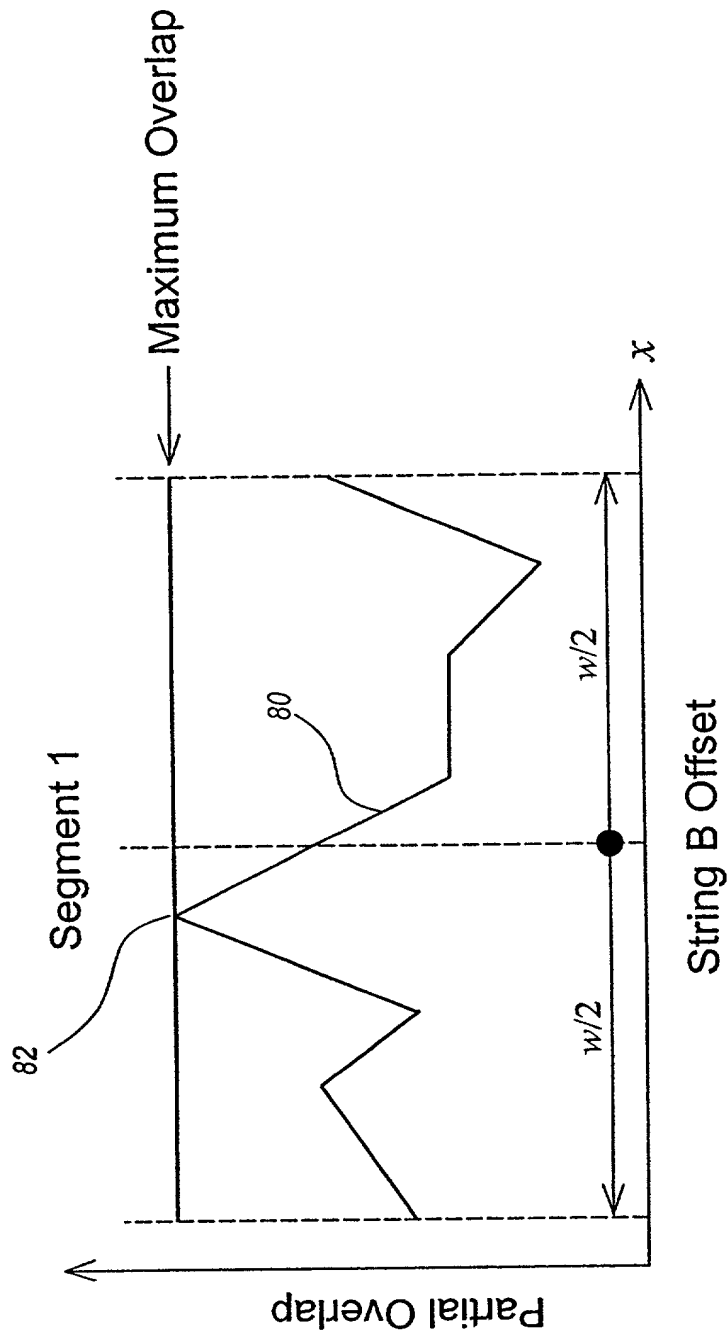


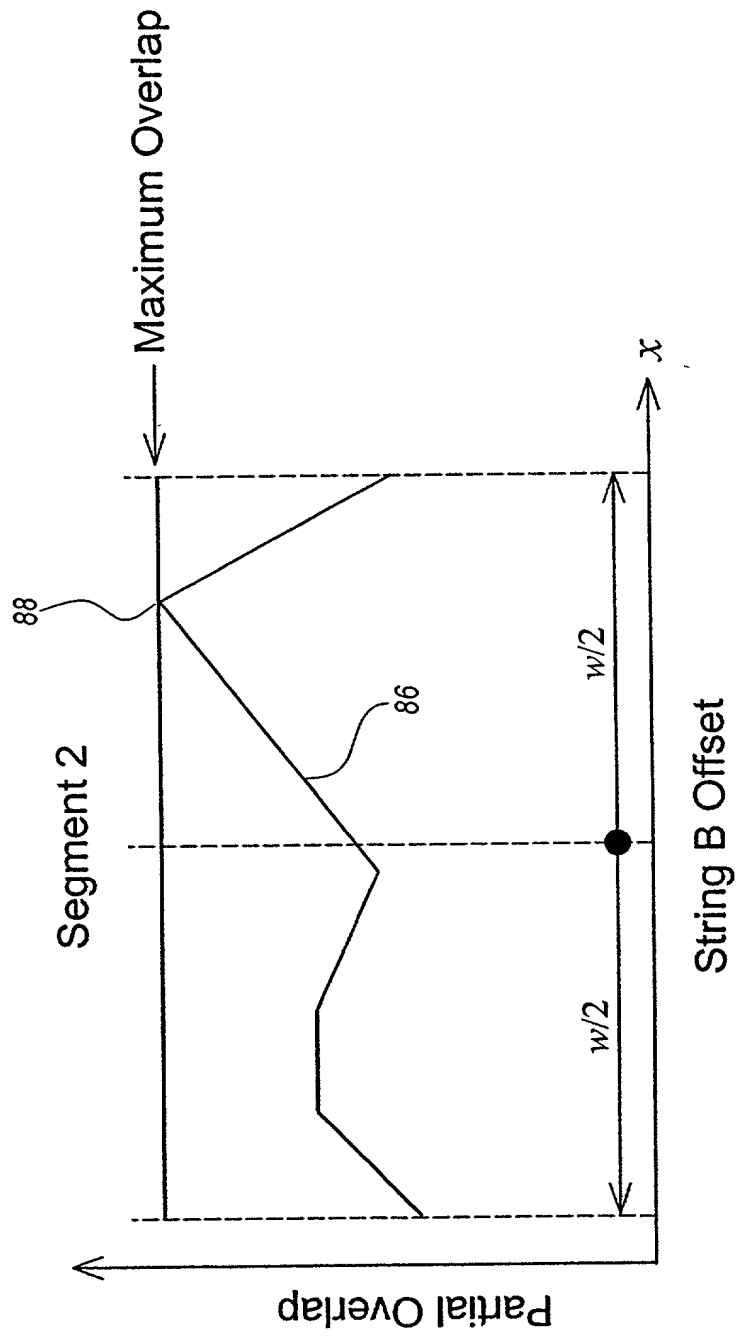
FIG. 6







**FIG. 8A**



**FIG. 8B**

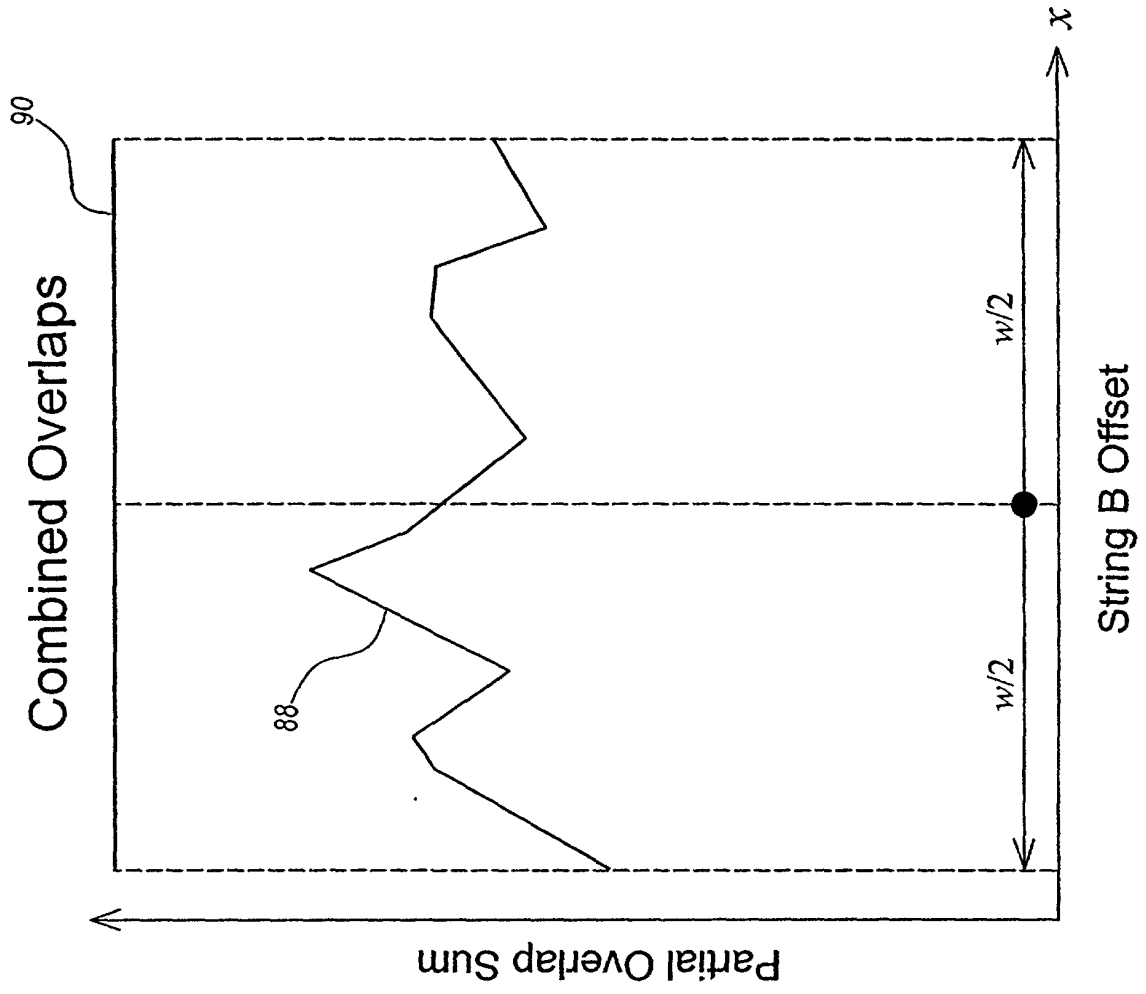
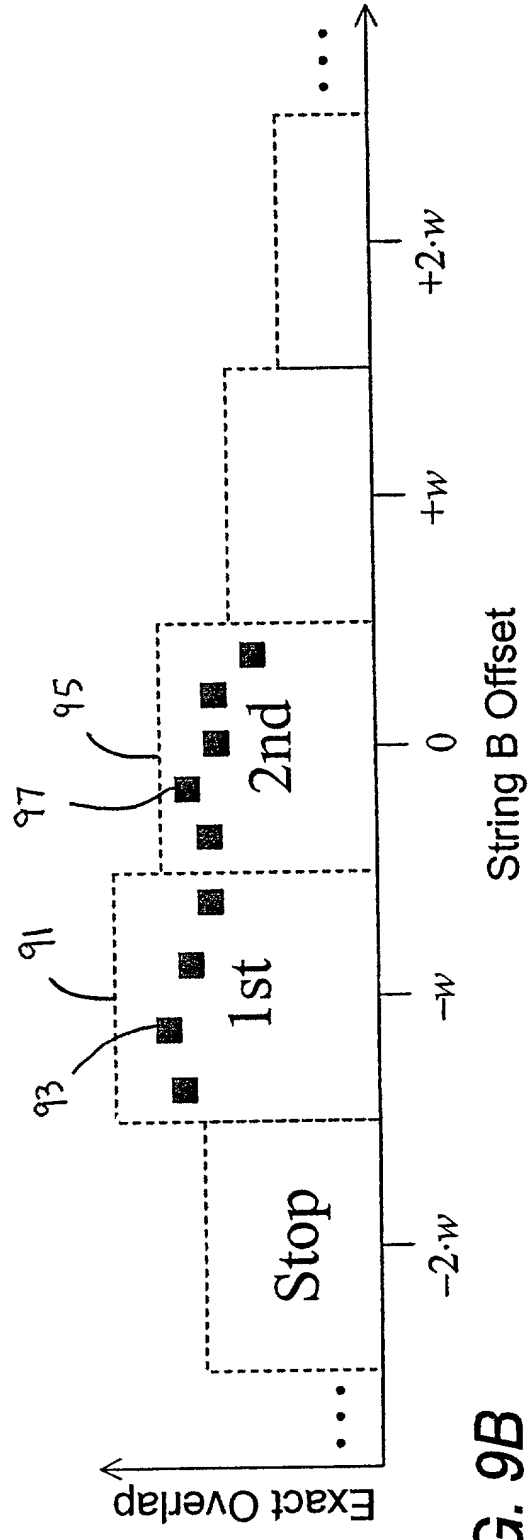
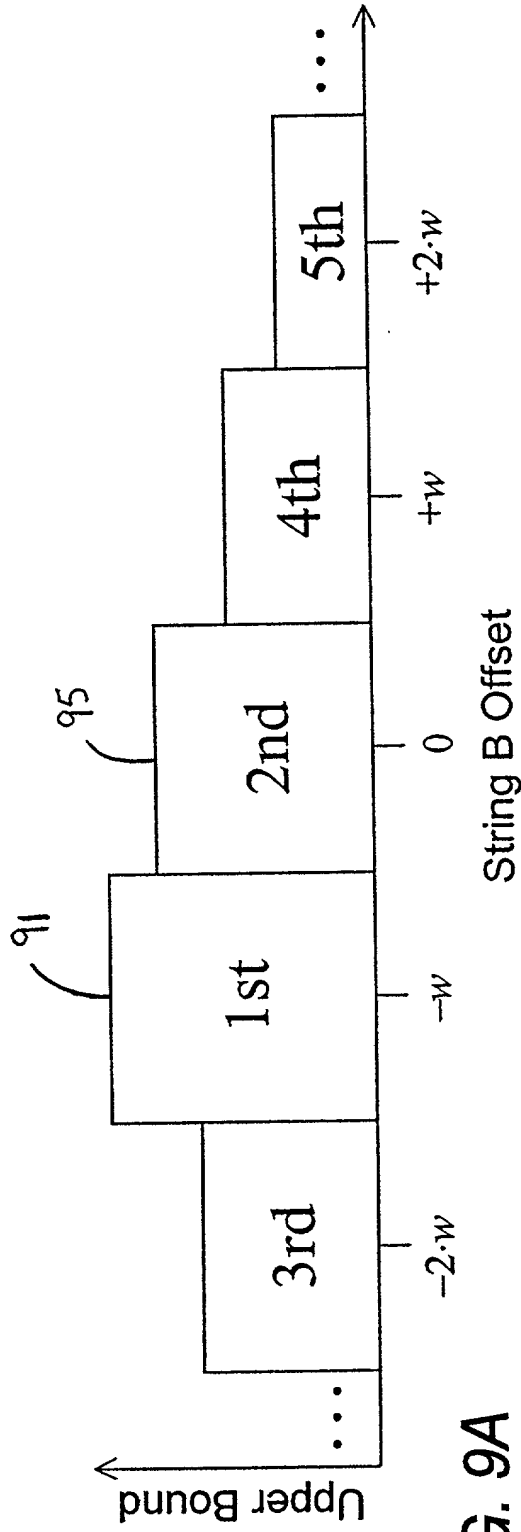
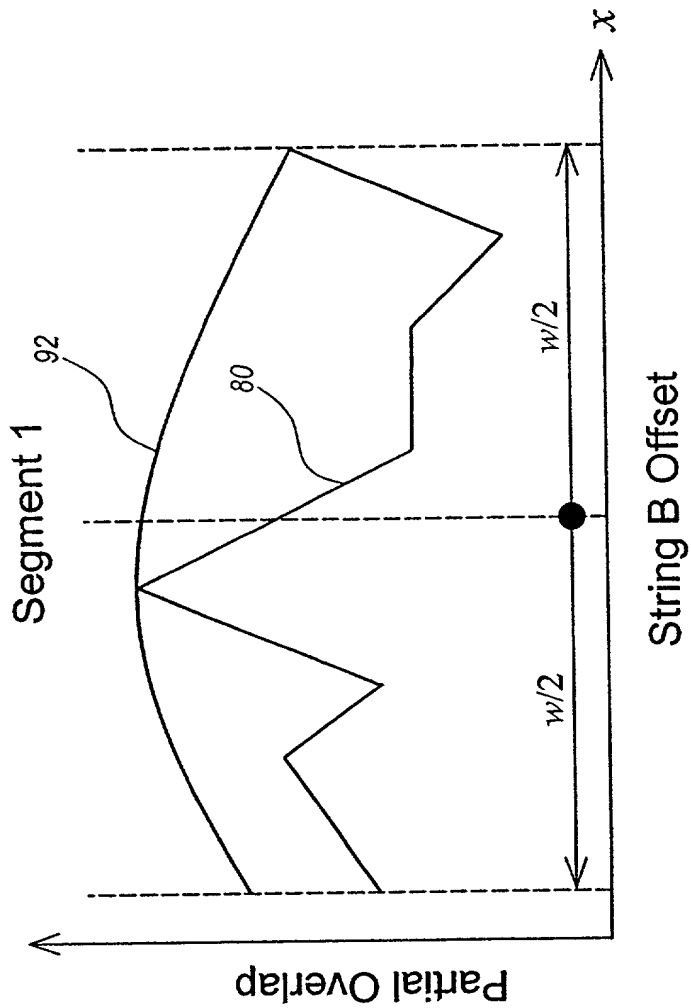


FIG. 8C





**FIG. 10A**

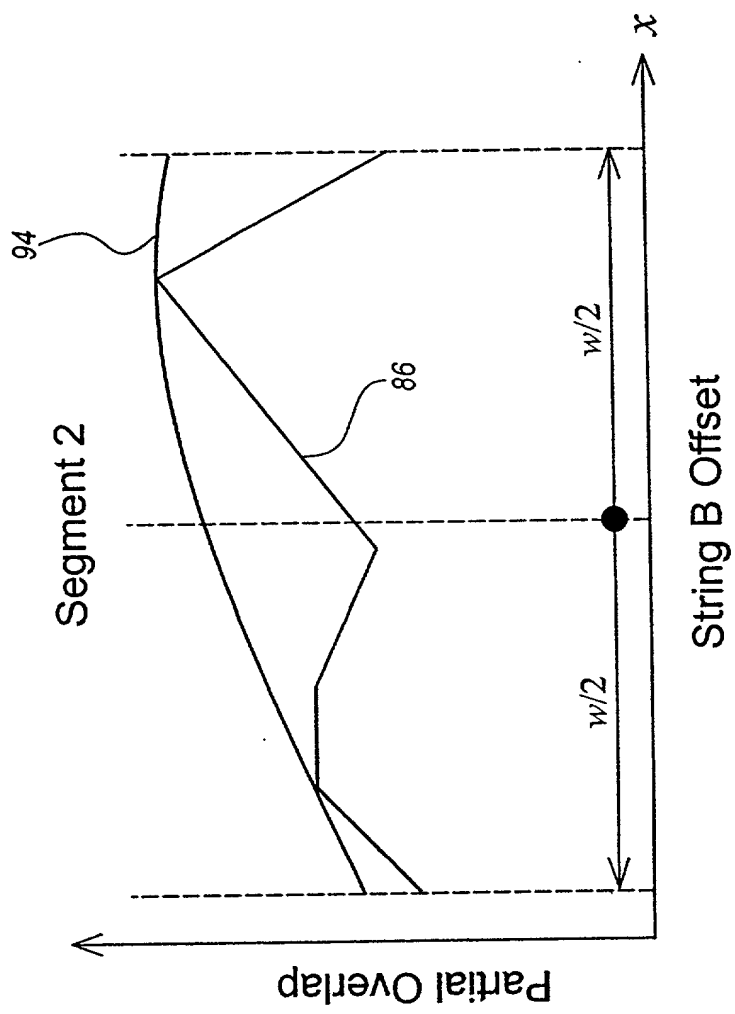
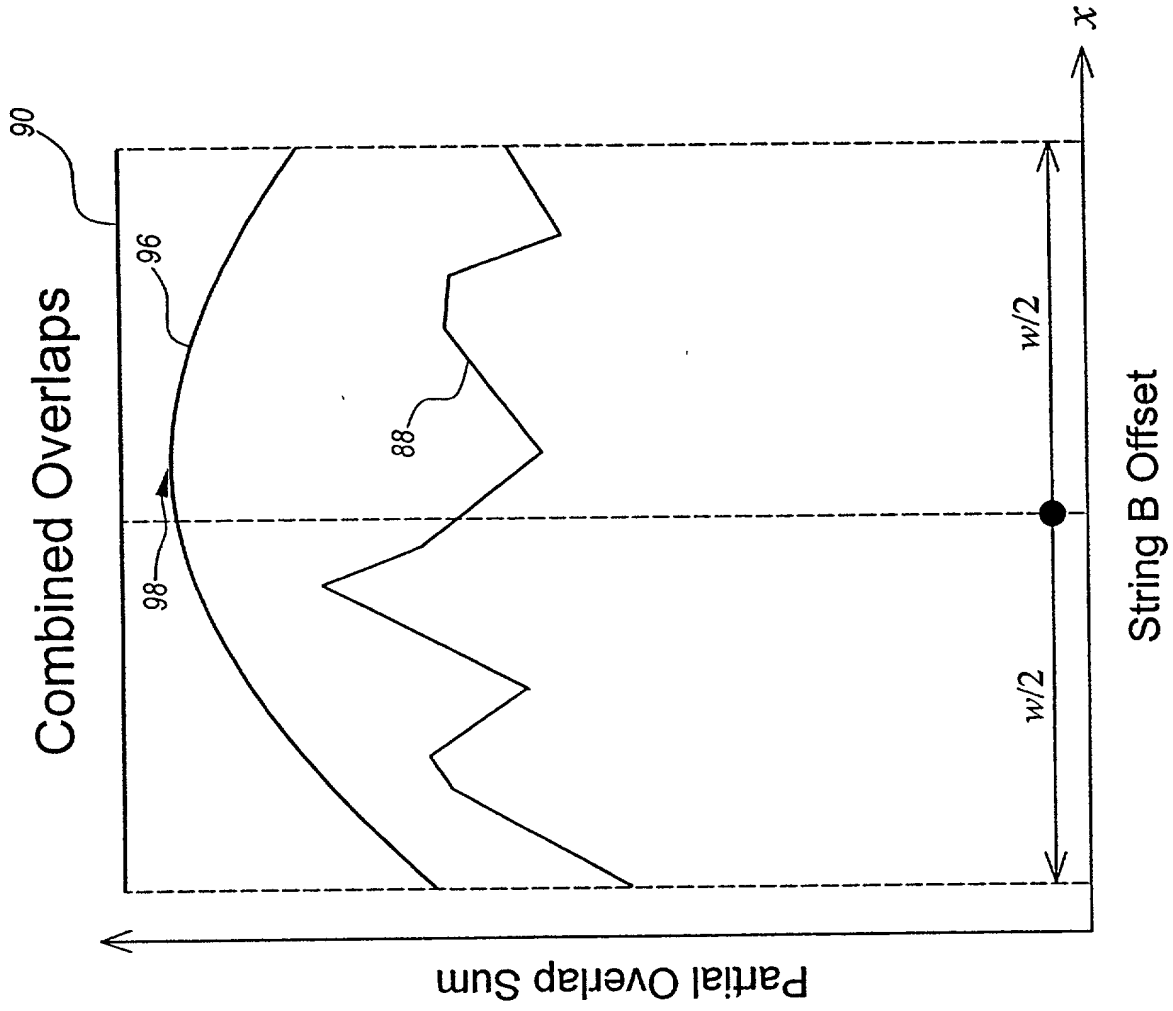
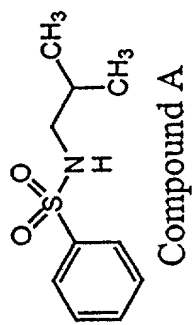


FIG. 10B



**FIG. 10C**





2D→1D

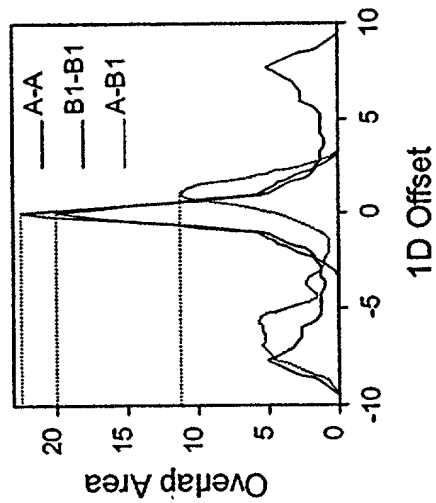
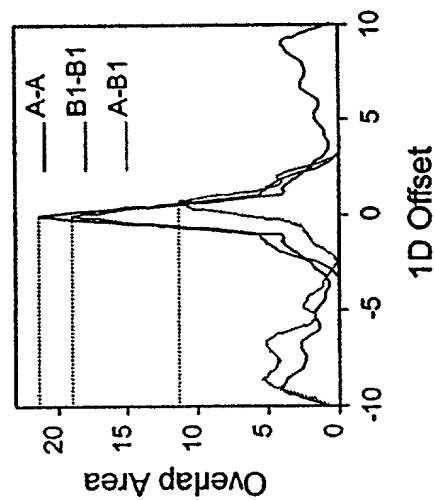
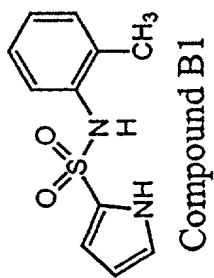
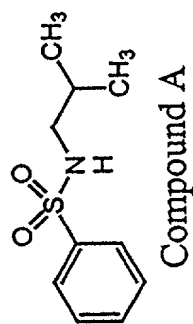
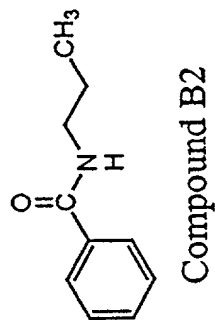


FIG. 11



2D→1D



3D→1D

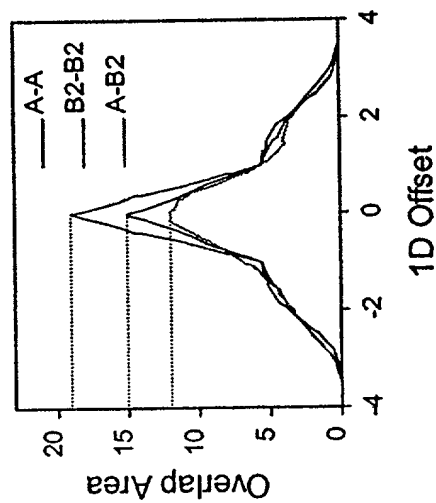
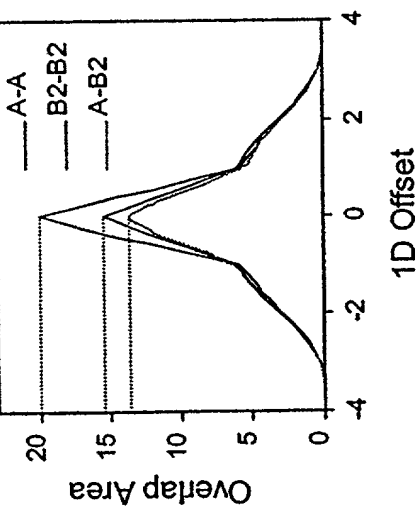
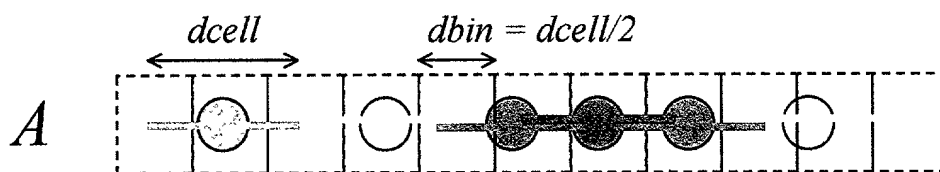






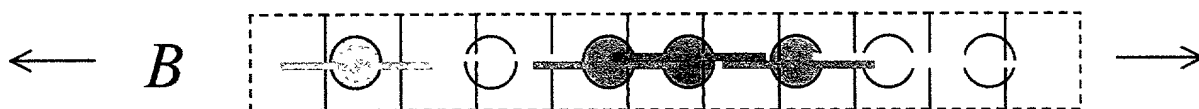
FIG. 12






# Bin-Based Overlap

- Do a series of fast overlap calculations using “bins” with integer occupation numbers (0→255) for each atom:



	170	255	25							
										
					204	255	51			
						179	255	76		

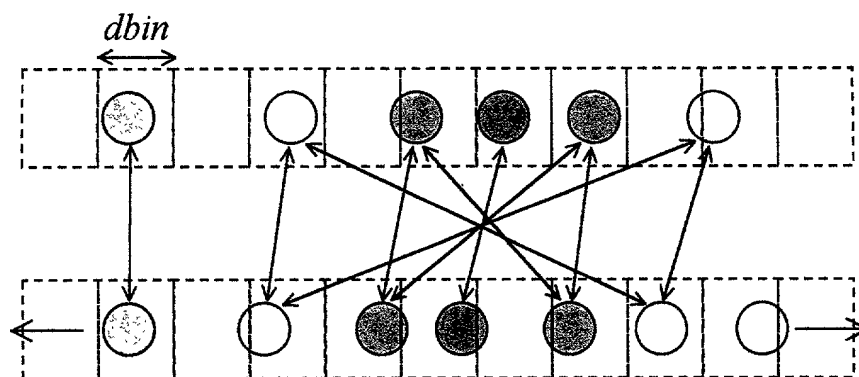


	170	255	25							
										
					48	255	207	209	255	46
						36	255	219		
										

- Multiply occupation numbers for matching atom types across aligned bins to get a good estimate of overlap area
- Fast, but there are numerous bin-based offsets that must be considered

FIG. 13

# Speeding Up Bin-Based Overlap Calculations



- 21 unique bin offsets, 10 matching atom type pairs
- There are only 6 different bin offsets wherein matching atom types are approximately aligned:

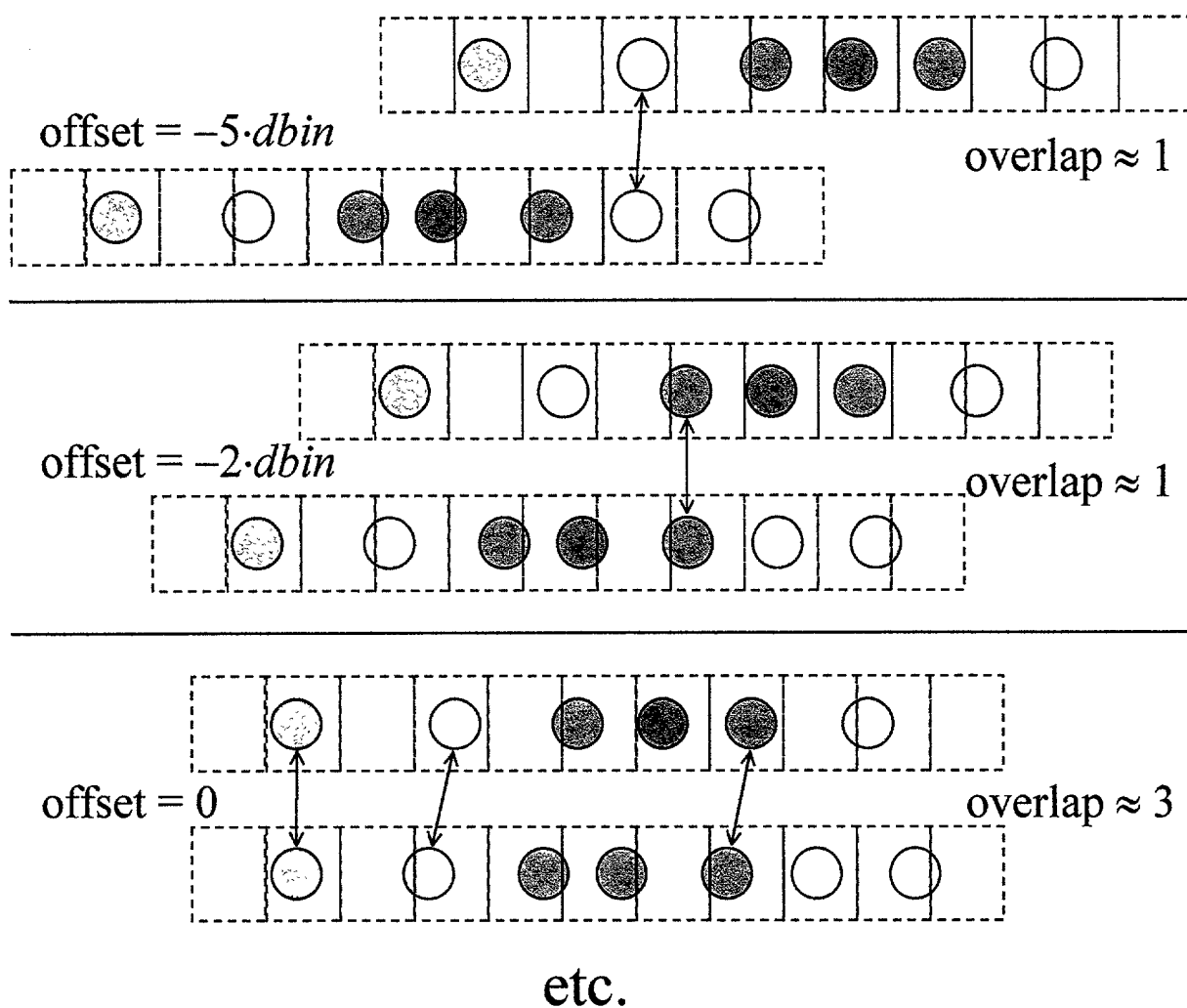
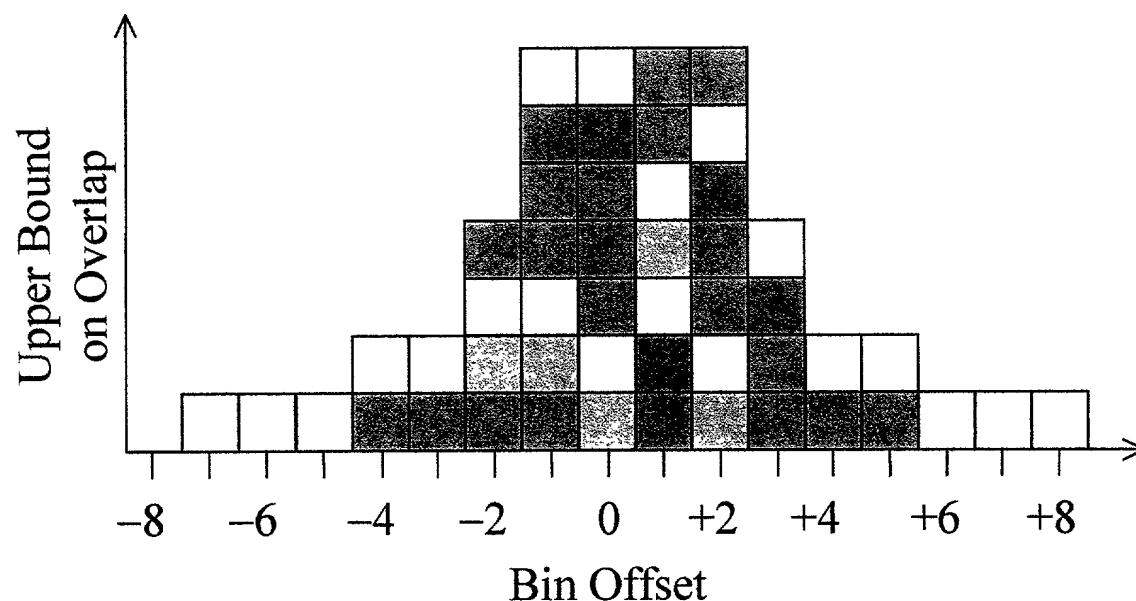
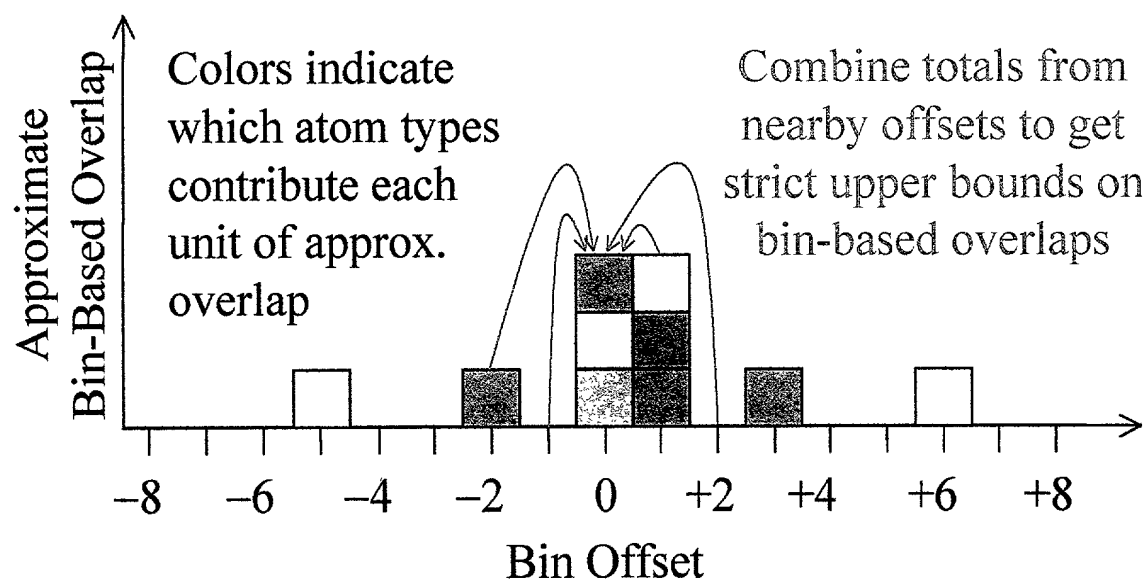


FIG. 14

## Approximate Bin-Based Overlaps → Upper Bounds



- Process offsets in order of decreasing upper bound
- Do standard bin-based overlap calculations (with occupation numbers), keeping track of the largest overlap value
- Stop when remaining upper bounds are lower than this largest bin-based overlap

FIG. 15